JEAS | Journal of Economics and Administrative Sciences
E-ISSN: 2148-1792 P- ISSN: 1302-2024
Volume 3, Supplement Issue 1
www.jeansweb.org

Business process mining in agriculture environment

Shabnam Navabi[*]
Master in IT Engineering, Agricultural Research, Education and Extension Organization, Tehran, Iran.

Corresponding Author

Fariba Mazraee
MSc in Management, Kerman, Iran.

Elaham Hamidi Motlagh
MSc in Management, Karaj, Iran.

ABSTRACT

Process mining is an emerging technology in the context of Business Process Management with the goal to derive process models from observed system behavior in available logging data. A process model is a graphical representation of a business process that describes the dependencies between activities (and constraints between them) that need to be executed collectively for realizing a specific business objective and can be used to reduce complexity by omitting unnecessary characteristics. Process mining is a relatively young research discipline that sits between computational intelligence and data mining on the one hand, and process modeling and analysis on the other hand. The idea of process mining is to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's (information) systems.

In this paper, we describe the application of process mining for the process of managing research proposals in IRAN Agricultural Research, Education and Extension Organization as the exclusive responsible of research in agriculture in country. For automating of this process, a Research Management System (RMS) is designed and implemented for managing the complicated process of offering until approve of research proposals at 2010. The system's event log is producing with programming since then. This system's log event has been surveyed during 5 years (2011-2015) and with 48381 records (only related to got approved proposal) and 1545 active users acting as 8 key roles. In our approach, we analyzed the process from three perspectives including control flow, organizational and performance perspectives. For mining we used Disco with its visualization capabilities and SQL query language. The results show model relative efficiency in detailed statistics.

*Keywords**:** Process Mining, Workflow Management, Business Process Management, Business Process Analysis*

Introduction

Process Mining, as experienced in data mining, can lead organizations Overcoming challenges posed by current processes, by discovering, monitoring and improving them, extracting knowledge from events log readily available in today's information systems.

Organizations expect process mining to produce accurate insights regarding their processes while depicting only the desired traits and removing all irrelevant details. In addition, they expect the results to be comprehensible and context-sensitive. Process Mining has the potential to extract management information by deriving process models from observed system behavior like system log files or produced log records by programming as our experienced method, will be described in this paper.

It assumes that it is possible to record events such that (i) each event refers to an activity (i.e., a well-defined step in the process), (ii) each event refers to a case (i.e., a process instance), (iii) each event can have a performer also referred to as originator (the actor executing or initiating the activity), and (iv) events have a timestamp and are totally ordered.

A major area of application for process mining is the discovery of formerly unknown process models for the purpose of analysis or optimization [5]. The process reengineering and the implementation of ERP systems in organizations gained strong attention starting in the 1990s. Practitioners have since primarily focused on designing and implementing processes and getting them to work.

With maturing integration of information systems into the execution of processes and the evolution of new technical possibilities the focus shifts to analysis and optimization.

Actual executions of business processes can now be described and be made explicit. The discovered processes can be analyzed for performance indicators like average processing time or costs for improving or reengineering the process. The major advantage of process mining is the fact that it uses reliable data. The date that is generated in the source systems is generally hard to manipulate by the average system user. For traditional process modeling necessary information is primarily gathered by interviewing, workshops or similar manual techniques that require the interaction of persons. This leaves room for interpretation and the tendency that ideal models are created based on often overly optimistic assumptions.

Analysis and optimization is not limited to post-runtime inspections. Instead it can be used for operational support by detecting traces being executed that do not follow the intended process model. It can also be used for predicting the behavior of traces under execution. An example for runtime analysis is the prediction of the expected completion time by comparing the instance under execution with similar already processed instances. Another feature can be the provision of recommendations to the user for selecting the next activities in the process. Process mining can also be used to derive information for the design of processes before they are implemented.

We have applied process mining techniques in AREEO. Due to importance and complexity of Research Projects approve process in AREEO typical data analysis is time consuming and tedious. Flexibility of approve manners caused process variations and high amount data in log events.

The significant of our research was our log events with highest precision in spite of practical experiences that reveals real-life logs. They are often far from ideal and their quality leaves much to be desired and most of them, tend to be incomplete, noisy, and imprecise. We present a detailed analysis and extracted results of 5 years' log events (2011-2015) accumulated at 'Research Projects Management System' with more than 4600 cases and 48000 activities.

After successful BI and OLAP experiences in our organization with 'Management Dashboard System' and managers welcomed, Process mining was considered as next step to improve of current process in this research wide-spreading set.

As beginning of this article we illustrate the basic concepts of process mining. We show how process mining can be used in special, flexible, exclusive scenarios, provide an overview of relevant tools to give an outlook of selected contemporary challenges and research questions with their answers. In section 2 we discuss about related work. Section 3 introduces AREEO as is. Section 4 describes our Research methodology including Creating and Filtering Events log, Disco capabilities and our challenges. Then section 5 presents mining the Process Perspective while section 6 is about organizational Process Perspective. Section 7 and 8 are about Mining the Performance Perspective and Mining the Case Perspective and finally Section 6 concludes the paper.

Related Work

In today's competitive environment, executive managers are struggling to reduce organizational cost while improving quality of processes. With this regards, understanding the internal process models and process improvement are vital. Lack of effective process improvement can lead to under performance results and dissatisfaction of customers. However, the characteristics of some processes make the process model discovery and improvement non trivial tasks. For example, healthcare processes are highly complex, dynamic, ad-hoc and many other disciplines are required to work together for having an insight into the processes (Rebuge and Ferreira, 2012). Business Process Analysis (BPA) [?] and in particular the concept of process mining (van der Aalst et al., 2003) can provide facilities for the aforementioned issues. Some analysis is possible using existing business process analysis techniques, however usual process analysis techniques in many environment are time consuming, costly and requires a lot of efforts for people to describe those [2], [4], [8], [12], van der Aalst et al., 2007.

Process mining aims to extract process knowledge from so-called "event logs", which may originate from all kinds of systems like enterprise information systems or hospital information systems [5]. Therefore, the goal of process mining is to discover, monitor and improve real process in various organizations. With this regards, three basic types of process mining are discovery, conformance and enhancement (Rozinat and van der Aalst, 2008). Event logs are the starting point for any process mining technique. Before any technique can be applied to the event log, information can directly be obtained from the log through the preprocessing step (Song et al., 2013). The idea of applying process mining in the context of workflow management was first introduced in (Agrawal et al., 1998).

So far, there have been many efforts on the discovery and conformance types and consequently many algorithms and techniques are available depending on the specific situation reported in the literature [?]. However, the enhancement and improvement of business processes have not been explored properly and remain open challenge. Through process mining, users can obtain business performance metrics, process models, organizational models, organizational relations, performance characteristics, etc. (van der Aalst et al., 2007, Song and van der Aalst, 2008, Maruster and Beest, 2009, Gűnther and van der Aalst, 2007). Recently, several organizations such as high-tech companies, hospitals, and municipalities utilize process mining techniques to improve their processes (Song et al., 2008, Mans et al., 2008, Rebuge and Ferreira, 2012, Reijers et al., 2009, Lemos et al., 2011, Rozinat et al., 2009, van der Aalst et al., 2007).

However, there is still a lack of proper studies applied process mining to even logs taken from real-life applications. Mining real-life case studies is very important and provides precious experiences. Existing tools, techniques and algorithms consider assumptions that do not necessarily take place in real situations (e.g. in dealing with exceptions and assuming there is no noise). Additionally, traditional process mining techniques produce valuable information in various perspectives when they are applied to well-structured processes, thus they generate lasagna-process models which are easy to understand (Jagadeesh Chandra Bose and van der Aalst, 2009, Gűnther and van der Aalst, 2007). However, lots of real-life business processes are unstructured processes, thus produce spaghetti-like process models which are difficult to understand. Real-life event logs are usually very huge and complicated, since the event logs contain numerous activities which are executed by many employees.

In this paper, we use such event log in AREEO to demonstrate many challenges and issues can occur in the organization. However, there is still a lack of case studies applied process mining in other domains that the community has to pay special attention in performing corresponding research.

The Agricultural Research, Education and Extension Organization (AREEO)

AREEO is the largest responsible body for agricultural research, education and extension in Iran. This organization has 25 specialized subject institutes. The establishment of some agricultural research institutes in Iran dates back to more than 90 years ago. Among these institutes, Razi Vaccine and Serum Research Institute (1924), Animal Science Research Institute (1933), Plant Pests and Diseases Research Institute (1943), and Seed and Plant Improvement Institute (1959) could be named as pioneer research institutes. More than 2100 scientific board members, working in these institutes and 31 provincial capitals, have

participated in more than 40000 Agricultural Research Projects for 40 years. On average, every year, about 2000 proposals have been offered and confirmed after complex and unique process. In Figure 1, this approve flow chart has been depicted.

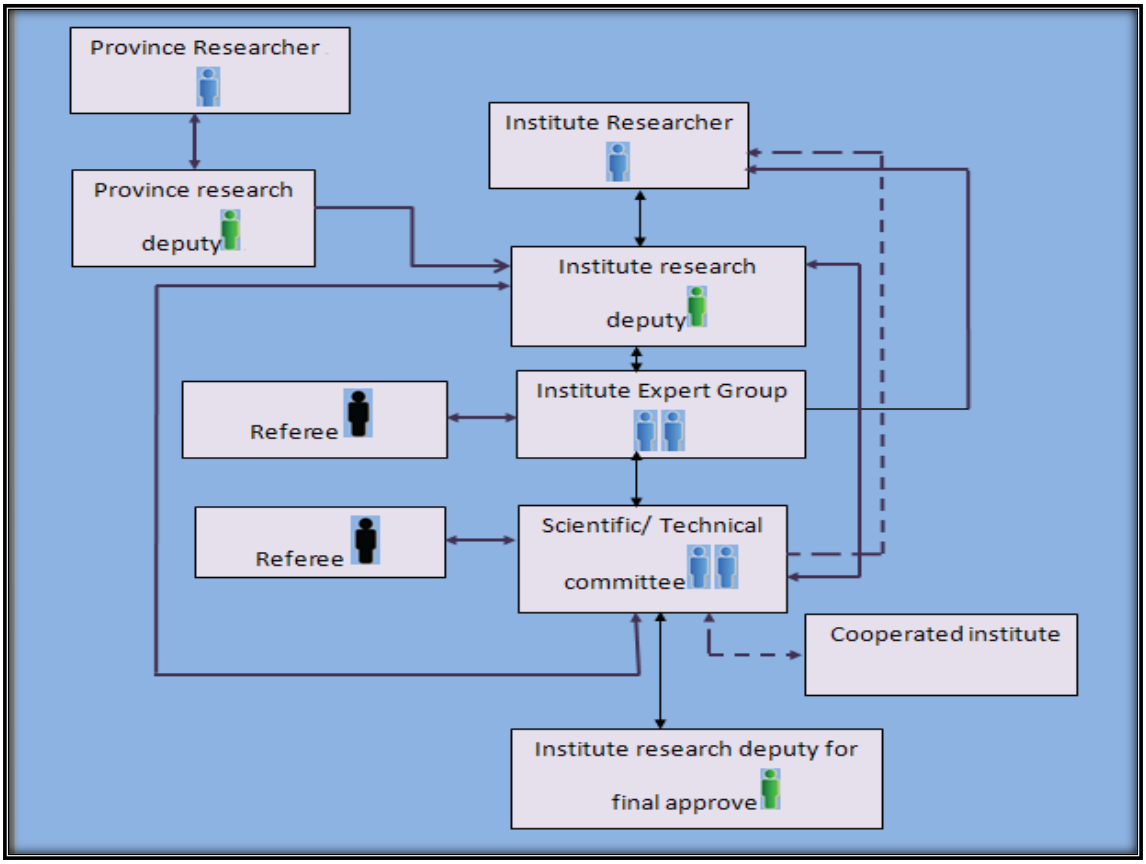


Figure 1: Process flow chart

All of highly specialized researchers (province researchers & institute researchers) are involved in the projects by field of agriculture in IRAN. They differ only in the place of employment. Definition of institute expert section or group is number of selected and appointed researcher with a common technical background and task to investigate and judge a proposal with the subject mentioned.

Scientific/Technical Committee has been established from number of faculty members with different specialized research field and is headed by Deputy of Research institute.

As shown in Figure 1, all of research proposal approved, send to Institute Expert Group. In this part, proposal passes the main stage of evaluation and verification. The Scientific/Technical Committee in institutes has the key role in final approve of a proposal. At last a proposal is approved, coded by institute research deputy and after that researcher commences his/her research.

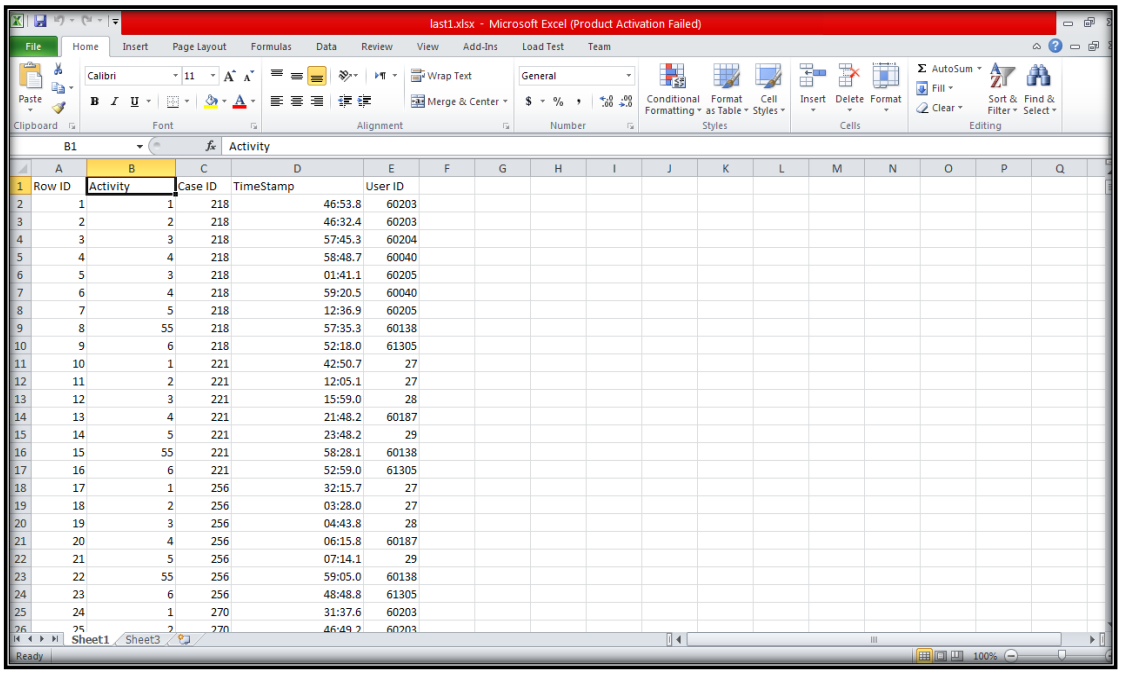
By observing the whole process model (Figure 1), it is obvious that discovering useful information or characteristics of activities is hard. With this regards, tracing of processes from the entire events log can be used to extract the most important information about the problematic activities or employees.

After mining the processes, events log can include many features which they are intended as a tool for management decision-making for process structure improvement and solving the problems.

The extent and distribution of AREEO increases the importance of improvement of process. Obviously, the managers of organization have recognized the importance of improving the structure of the most critical and longest process as well as the answers to their questions that tracing of process can give them.

As indicated in the introduction, the basis for all process mining techniques is a process log. Such a log is a file generated by some information system, with information about the execution of a process. Since each information system has its own format for storing log files, we have developed an Excel File from our SQL Server DB for Disco framework to store a log in. For clarity and to avoid complexity, in this paper, we decided Proposals before approval have been filtered. So, the excel file includes more than 48000 records of events log containing columns of case ID, Activity ID, Timestamp and user ID (The ID of user as the initiator of the event).

The 'Research Projects Management System' has been developed that creates log events automatically, as records of SQL Server DB, such that (i) Every row corresponds to one event (ii) each event refers to an activity (i.e., a well-defined step in the process), (iii) each event refers to a case (i.e., a process instance), (iv) each event can have a performer also referred to as originator (the actor executing or initiating the activity), and (v) events have a timestamp and are totally ordered. You can see information about Case IDs, Activities IDs, Start time, Resources in Figure 2.



Row ID	Activity	Case ID	TimeStamp	User ID
1	1	218	46:53.8	60203
2	2	218	46:32.4	60203
3	3	218	57:45.3	60204
4	4	218	58:48.7	60040
5	3	218	01:41.1	60205
6	4	218	59:20.5	60040
7	5	218	12:36.9	60205
8	55	218	57:35.3	60138
9	6	218	52:18.0	61305
10	1	221	42:50.7	27
11	2	221	12:05.1	27
12	3	221	15:59.0	28
13	4	221	21:48.2	60187
14	5	221	23:48.2	29
15	55	221	58:28.1	60138
16	6	221	52:59.0	61305
17	1	256	32:15.7	27
18	2	256	03:28.0	27
19	3	256	04:43.8	28
20	4	256	06:15.8	60187
21	5	256	07:14.1	29
22	55	256	59:05.0	60138
23	6	256	48:48.8	61305
24	1	270	31:37.6	60203
25	2	270	46:49.2	60203

Figure 2: Event logs in Excel format

Our knowledge about our events log extracted by SQL Queries is:

Table 1: Events log features

How many cases (or process instances) are in the log?	4613
How many tasks (or audit trail entries) are in the log?	48381
How many resources are in the log?	1545
How many roles are in the log?	8
Are there running cases in the log?	NO

Research Methodology

In this section, we describe how Process Mining answered to and solved a lot of challenges in our main organizational process. Because our case is a process developed by SQL Server and .Net C#, we had this option that programming and queries can be used as potential capable method for mining. We used Disco 1.9.1 tool to develop our experiments.

- **Creating and Filtering Events log**

Our goal of this research is answering to the basic management questions about system after 5 years' experience. The first step is extracting the real map of current process. Disco tool gets a proper Excel file as input. In 'Research project management system', events log is produced programmatically (by programming). Each record has fields as id, id_proposal, erjadate, id_user and activty_id. In every transaction which occurs by users in system, one record adds to events log. We export this event log to Excel in SQL Server. The first step in Disco, is matching existing column with defined column. It is obvious that this step passes easily because of common definition in both systems. After the first process in Disco, some noisy processes were found and deleted from log. Figure 3 shows the output map before ignoring noisy process. One of the most applicable features in Disco is filtering one special rout in the map and then showing the related cases. We used this feature to clear our event log from noisy data. More than 500 records were found and deleted in this step. (SQL queries and Disco have been used.)

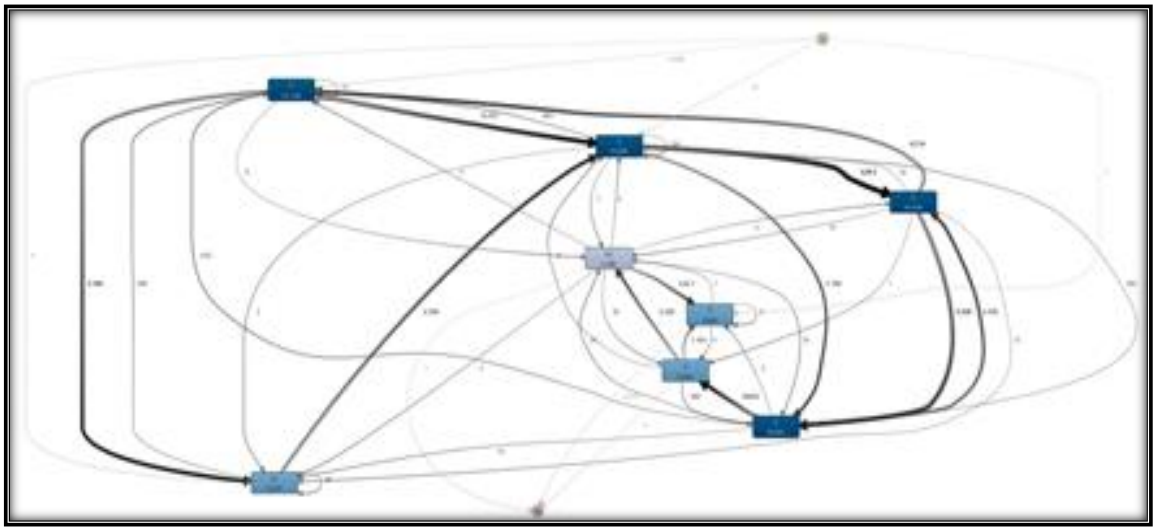


Figure 3: Disco output map from events log before filtering

- **Analysis Plan**

Process mining can analyze your process in a bottom-up fashion. You don't need to have a model of your process to analyze it. Process mining uses the history data in your IT systems. Your IT system already records all steps of your process in execution. With process mining, you get a process model from these data. This way, your real process, and actual business rules, can be discovered automatically. We already have a model of how your process should be performed? With process mining, we get objective information on whether it is actually followed as prescribed.

We may have a process that is well-defined, but not tightly enforced by your IT systems. Now we can see for the first time how that process is handled in real life. (www.fluxicon.com)

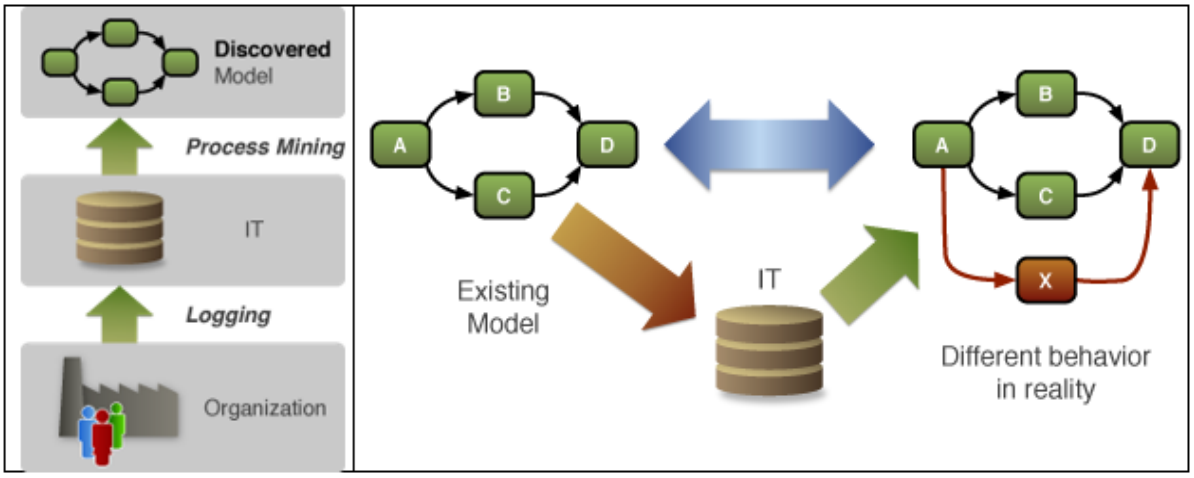


Figure 4: Analysis Plan

- **Why Disco?**

The process mining technology in Disco can automatically create smart flow diagrams of our process. All we need are event logs that are already on our IT systems. Because Disco works with this objective information, we no longer need to rely on belief or hearsay. Disco easily overcame on our unknown obstacles in mining the process. Disco was handcrafted by experts. Disco was built by former leading academics with more than eight years of process mining experience. Disco is the result of experience from countless process mining projects and was designed as a tool that fits perfectly into the workflow of professionals. Disco was optimized for speed. Disco contains the fastest process mining algorithms, and the most efficient log management and filtering framework. On top of that, we can get an obsessively streamlined user experience, allowing us to move fast. Disco was designed for human. Software should serve the user, not the other way around. We are already a process improvement expert, and we shouldn't need to become an expert in using process mining software. With Disco, we don't need to, because it is intelligent, sane, and fun.

Disco is a great process mining tool that simply works: it is able to deal with large event logs and complex models and conversion and filtering are made easy. Performance metrics are shown in a direct and intuitive manner and the history can be animated on the model, Process mining for the masses. Disco allows us to focus on the job at hand rather than the tool. (www.fluxicon.com)

- **Challenges**

Due to the variety of paths that can be taken, the first question we should answer was the ideal model to approve a proposal. In our case, this model corresponds with answer of this question: What is the most frequent path for the process model? We find this simple model, showed in Figure4, for approve a research proposal, by a query in our DB, will be shown later. It starts with the propose of a project and goes on with evaluating and approving by the Institute Deputy Research, Expert Group and Scientific/Technical Committee and ends with final approval and coding by Institute Deputy Research.

SQL, is a very capable feature, enables programmer to extract answers to any kind of questions from data gathered in relational databases. In our case we used SQL Server 2012 as the database. One of the questions that be answered by SQL queries was: What is the most frequent path for the process model?

For answering this question, by considering the event logs showed in Figure 2, at first we create a new Table with these codes in C# and ASP.Net:

Protected void change_click(object sender, EventArgs e)

```
Protected void change_click(object sender, EventArgs e)
{

String sql;
```

```
sql = "SELECT * FROM Eventlog  order by id_proposal,id ";
DataSet ds = new DataSet();
ds = ExecuteSqlBySDA(sql);
int temp = 0;
int r = 1;
int j = 0;
while (j <ds.Tables[0].Rows.Count)
{
       temp = Convert.ToInt32(ds.Tables[0].Rows[j]["id_proposal"]);
       string sql1;
       sql1 = "SELECT * FROM Eventlog where id_proposal=" + temp + " order by id ";
       DataSet ds1 = newDataSet();
       ds1 = ExecuteSqlBySDA(sql1);

       for (int i = 0; i < ds1.Tables[0].Rows.Count;i++)
       {
              DateTime s1;
              DateTime s2;
              TimeSpants;
              Int differenceInSeconds;

              if (Convert.ToInt32(i + 1) < ds1.Tables[0].Rows.Count)
              {
                     s1 = Convert.ToDateTime(ds1.Tables[0].Rows[i]["ErjaDate"]);
                     s2 = Convert.ToDateTime(ds1.Tables[0].Rows[i + 1]["ErjaDate"]);
                     s = s2 - s1;
                     differenceInSeconds = ts.Days;
                     // Difference in Seconds.
                     sql = "update [transactiondelays] set [F" + r + "] = " +
                     Convert.ToString(ds1.Tables[0].Rows[i]["place"]) + ",[F" + r + "F" +
                     Convert.ToInt32(r + 1) + "]= " + differenceInSeconds + " where
                     id_proposal=" + temp;
              }
              else
              {
                     sql = "update [transactiondelays] set [F" + r + "] = " +
                     Convert.ToString(ds1.Tables[0].Rows[i]["place"]) + " where
                     id_proposal=" + temp;
              }
              ExecuteSql(sql);


              r = r + 1;

       }
j = j + 1;
r = 1;

}

}
```

This new table has the following format: id_proposal and F1, F2…F60 which consequently show the passed steps by every proposal. This step has showed with numbers.

Figure 5: The extracted Table by coding as the steps

id_proposal	F1	F2	F3	F4	F5	F6	F7	F8	F9
218	1	2	3	4	3	4	5	55	6
221	1	2	3	4	5	55	6		
256	1	2	3	4	5	55	6		
270	1	2	3	4	5	55	6		
300	1	2	3	4	5	55	6		
309	1	2	3	4	5	55	6		
337	1	2	3	4	5	55	6		
339	1	2	3	4	5	6			
342	1	2	3	4	5	55	6		
377	1	2	3	4	5	55	6		
382	1	2	3	4	5	55	6		
393	1	2	3	4	5	55	6		

This table has a row for every proposal which show its path. By grouping and counting similar paths the question can be answered easily with this SQL query:

```
SELECT  COUNT(id_proposal) AS count, F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F11, F12,
F13, F14, F15, F16, F17, F18, F19, F20, F21, F22, F23, F24, F25, F26, F27, F28, F29, F30,
F31, F32, F33, F34, F35, F36, F37, F38,
                F39, F40, F41, F42, F43, F44, F45, F46, F47, F48, F49, F50, F51, F52, F53,
F54, F55, F56, F57, F58, F59, F60
FROM    [transaction]
WHERE  (F1 = 6) OR (F2 = 6) OR (F3 = 6) OR (F4 = 6) OR(F5 = 6) OR (F6 = 6) OR (F7 =
6) OR (F8 = 6) OR (F9 = 6) OR (10 = 6) OR
 (F11 = 6) OR (F12 = 6) OR (F13 = 6) OR (F14 = 6) OR (F15 = 6) OR (F16 = 6) OR (F17 =
6) OR (F18 = 6) OR (F19 = 6) OR (F20 = 6) OR
 (F21 = 6) OR (F22 = 6) OR (F23 = 6) OR (F24 = 6) OR (F25 = 6) OR (F26 = 6) OR (F27 =
6) OR (F28 = 6) OR (F29 = 6) OR (F30 = 6) OR
 (F31 = 6) OR (F32 = 6) OR (F33 = 6) OR (F34 = 6) OR (F35 = 6) OR (F36 = 6) OR (F37 =
6) OR (F38 = 6) OR (F39 = 6) OR (F40 = 6)
GROUP BY F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16, F17, F18,
F19, F20, F21, F22, F23, F24, F25, F26, F27, F28, F29, F30, F31, F32, F33, F34, F35, F36,
F37, F38, F39, F40, F41, F42, F43, F44, F45, F46, F47, F48, F49, F50, F51, F52, F53, F54,
F55, F56, F57, F58, F59, F60
ORDER BY expr1 DESC
```

Figure 6: SQL Query for grouping paths

The result of query is:

Figure 7: SQL Query answer of grouping paths

538	1	2	4	5	55	6			
349	1	2	3	4	5	55	6		
146	1	2	4	5	6				
99	1	2	3	4	5	6			
91	1	11	2	3	4	5	55	6	

69	1	2	4	5	4	5	55	6	
60	1	2	3	4	3	1	2	3	4
58	1	2	3	4	3	1	2	4	5
53	1	11	2	3	4	3	1	11	2
51	1	2	3	4	3	1	2	3	4
48	1	2	3	1	2	3	4	5	6
47	1	11	2	4	5	55	6		
45	1	2	3	4	5	4	5	55	6
44	1	2	3	4	3	4	5	55	6
38	1	11	2	3	1	11	2	3	4
35	1	11	2	3	4	5	6		
34	1	2	1	2	4	5	55	6	
32	1	2	3	4	3	1	2	3	4
29	1	2	3	4	3	1	2	4	3
26	1	11	2	3	4	3	1	11	2

And the answer of question is:

Figure 8: The ideal proposal approve model for institute researcher

The model, based on mentioned kinds of researcher, changes to Figure9:

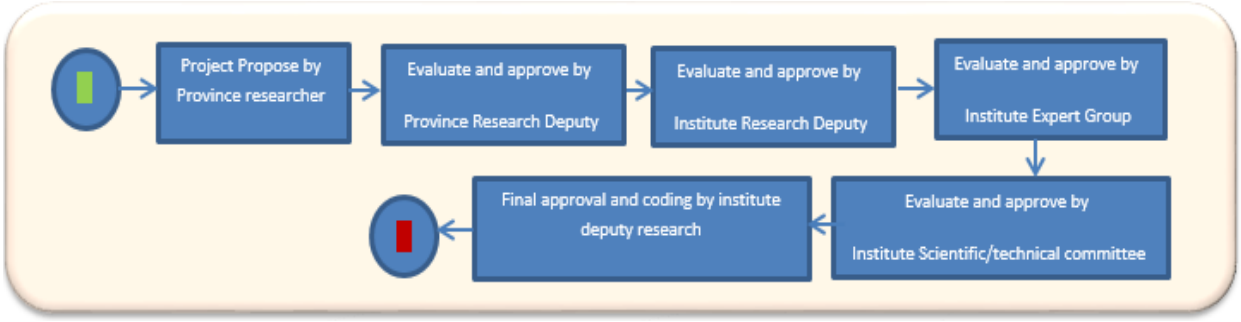


Figure 9: The ideal proposal approve model for province researcher

Another simple way is when institute decides to ignore the step of 'Evaluation and approve by Institute Expert Group', showed in Figure 10.

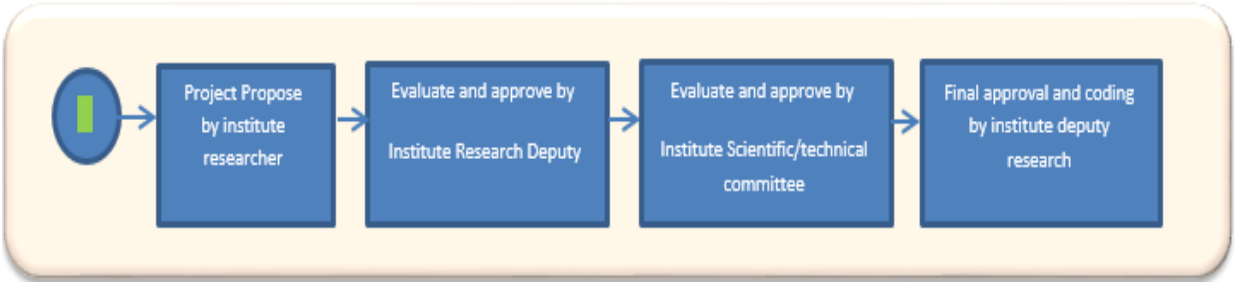


Figure 10: The ideal proposal approve model for institute researcher (Ignores one step)

The other Questions that managers usually have about performance of processes in organizations are: What is the average/minimum/maximum throughput time of cases? Which paths take too much time on average? How many cases follow these routings? What are the critical sub-paths for these paths? What is the average service time for each task? How much time was spent between any two tasks in the process model?

Are the rules indeed being obeyed? How many people are involved in a case? What is the communication structure and dependencies among people? How many transfers happen from one role to another role? Who are important people in the communication flow? (The most frequent flow) Who subcontract work to whom? Who work on the same tasks?

For analyzing process and answer to such questions, we chose Disco.

Disco framework shows the time performance as well as answers to mentioned questions.

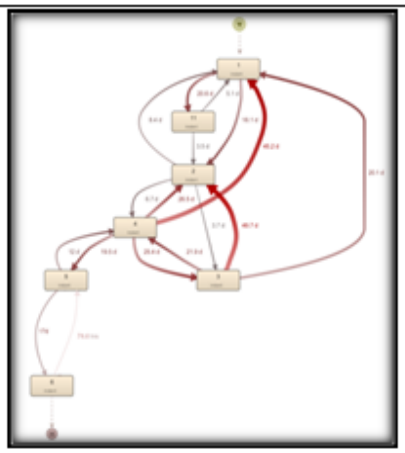


Figure 11: Mean Duration time of events

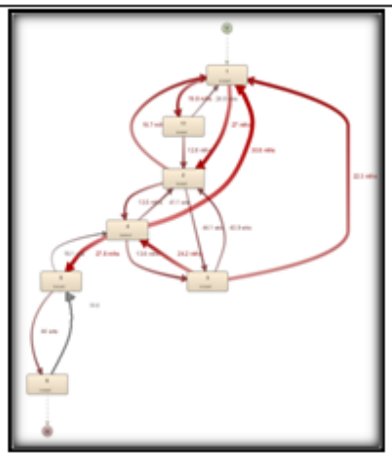


Figure 12:Max Duration time of cases

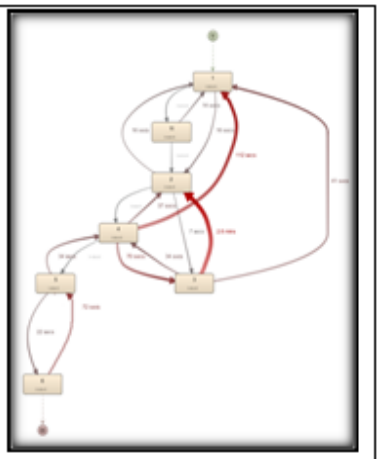


Figure 13:Min Duration time of cases

Mining the Process Perspective

This question that either the developed process has enough efficiency or not, comprising more expert studies but the output of this survey, shows this process benefits of high flexibility and can be claimed the most effort has been made for users' satisfaction.

Figure 14 and 15 are separated because of start point (Institute Researcher and Province Researcher); show all of passed paths, by two parameters (absolute frequency and mean duration) in 2011-2016. Figure 14 shows the process of evaluation and approve of 3049 proposals, registered by Institute Researcher. 2178 cases approved by former method in headquarter department and two steps (check and approve) and 871 cases approved after assignment of approve to Institutes. The lowest latency as shown in Diagram1 is between Researcher and Institute researcher; takes time 6.4 days and the most latency (71.7 days) appertain

to transactions between the roles of Institute Expert Group and Researcher. Although it causes of the key role of Institute Expert Group in evaluation, judgment and determination of reform comments about proposal. Figure 15 shows the process of evaluation and approve of 1563 proposals, registered by Province Researcher. 1080 cases approved by former method in headquarter department and two steps (check and approve) and 483 cases approved after assignment of approve to Institutes. The lowest latency as shown in Diagram2 is between Province Research Deputy and Institute Research Deputy; takes time 4.1 days and the most latency (70.5 days) appertain to transactions between the roles of Institute Expert Group and Researcher. Comparison of proposed research subject by Province Researcher (563 cases) and Institute Researcher (3049 cases) can be significant factor and impacts to management decision support essentially. On the other hand, 8% of Province proposals returned to Researcher for edition after evaluation by Institute Expert Group, while this number is 5.3 % for Institute proposals. This comparing is more notable about 84% returned proposals by Institute Scientific/Technical Committee to Province Researcher and 44.2 % to Institute Researcher.

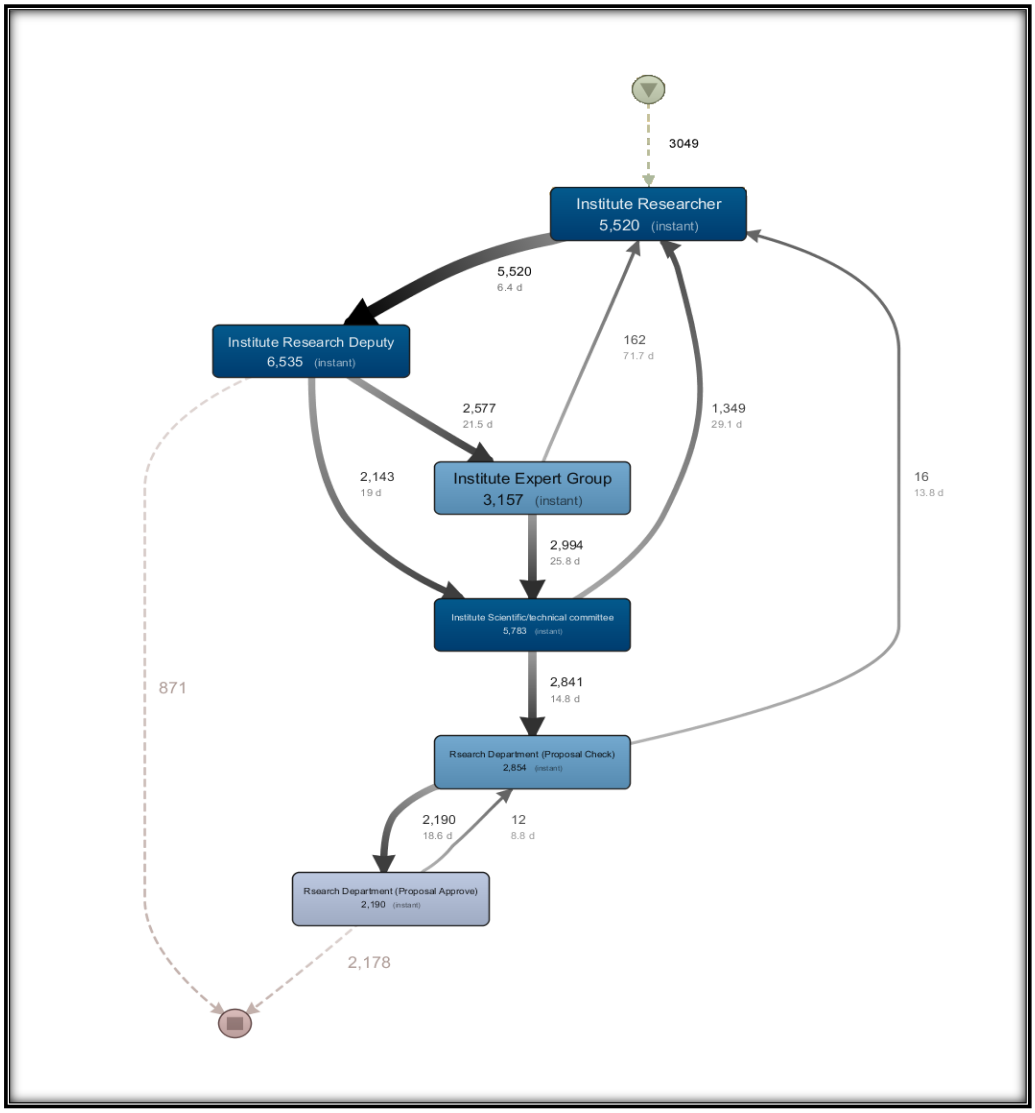


Figure 14: process perspective

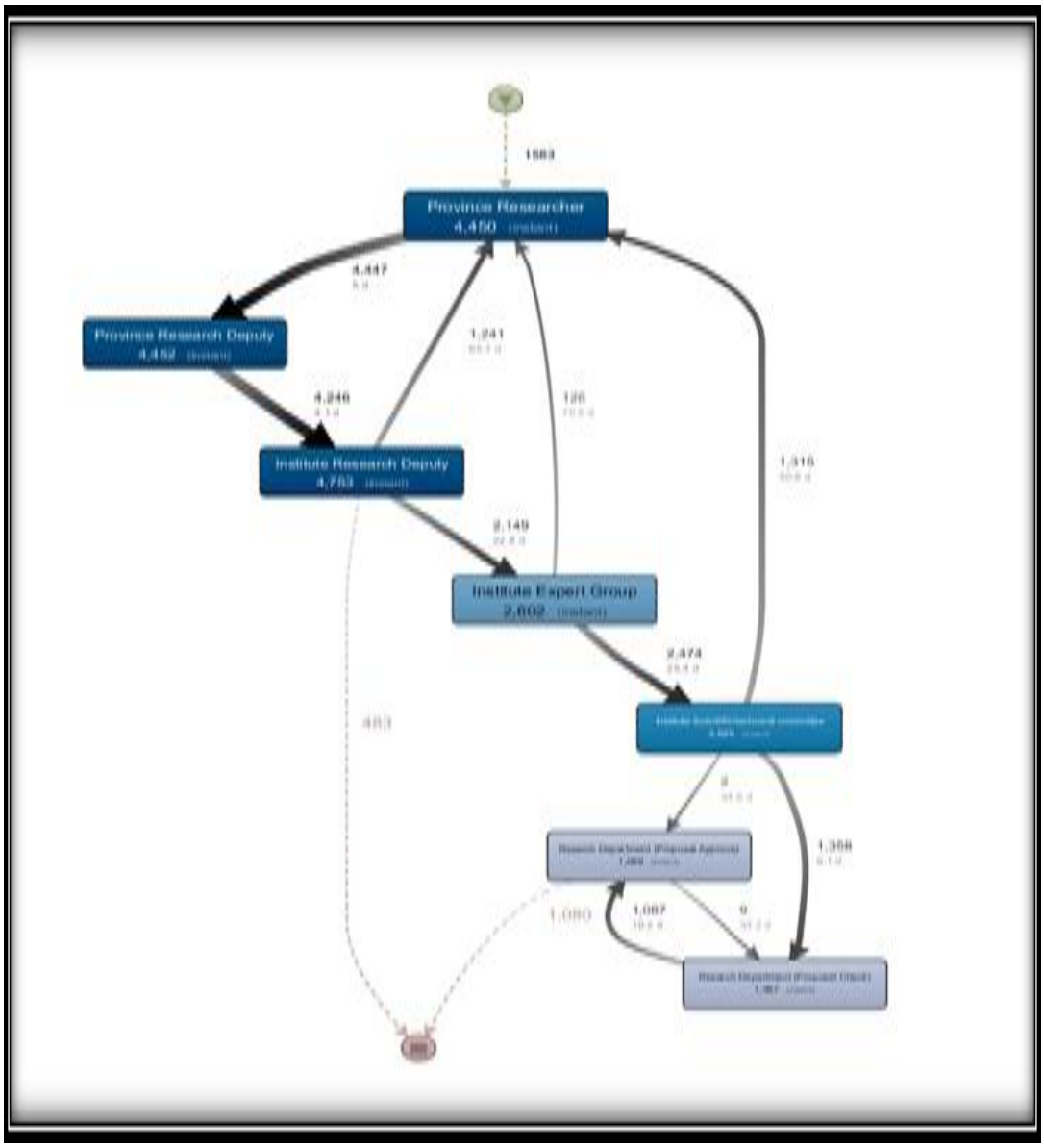


Figure 15: process perspective

Mining the Organizational Perspective

As explained, this system has 8 basic roles. Of course, after management changes in Approve Process, the duty of two headquarter role, Research Department (Check) and Research Department (Approve), has assigned to Institute research deputy.

Figure 16 shows the participation rate of Roles in the whole of process.

Figure 16: participation rate of Roles in the whole of process

Activity	Frequency	Relative frequency
Institute Research Deputy	11,288	23.33%
Institute Scientific/technical committee	9,412	19.45%
Institute Expert Group	5,759	11.90%
Institute Researcher	5,520	11.41%

Province Research Deputy	4,452	9.20%
Province Researcher	4,450	9.20%
Research Department (Proposal Check)	4,221	8.72%
Research Department (Proposal Approve)	3,279	6.78%

Obviously, because of multiplicity in role assigned to Institute research deputy, it has the most frequencies (23.3%) in reference. The next key role is Institute Scientific/technical committee which has last position as survey point to evaluate a proposal and 20% of confirm or reject transactions done by it.

It is notable that this committee has this potency that reject proposal to 1) Researchers, 2) Institute Research Deputy and 3) Institute Expert Group discretionally. The following Figure shows the average time that proposals for review or verification by any of the above roles are spent.

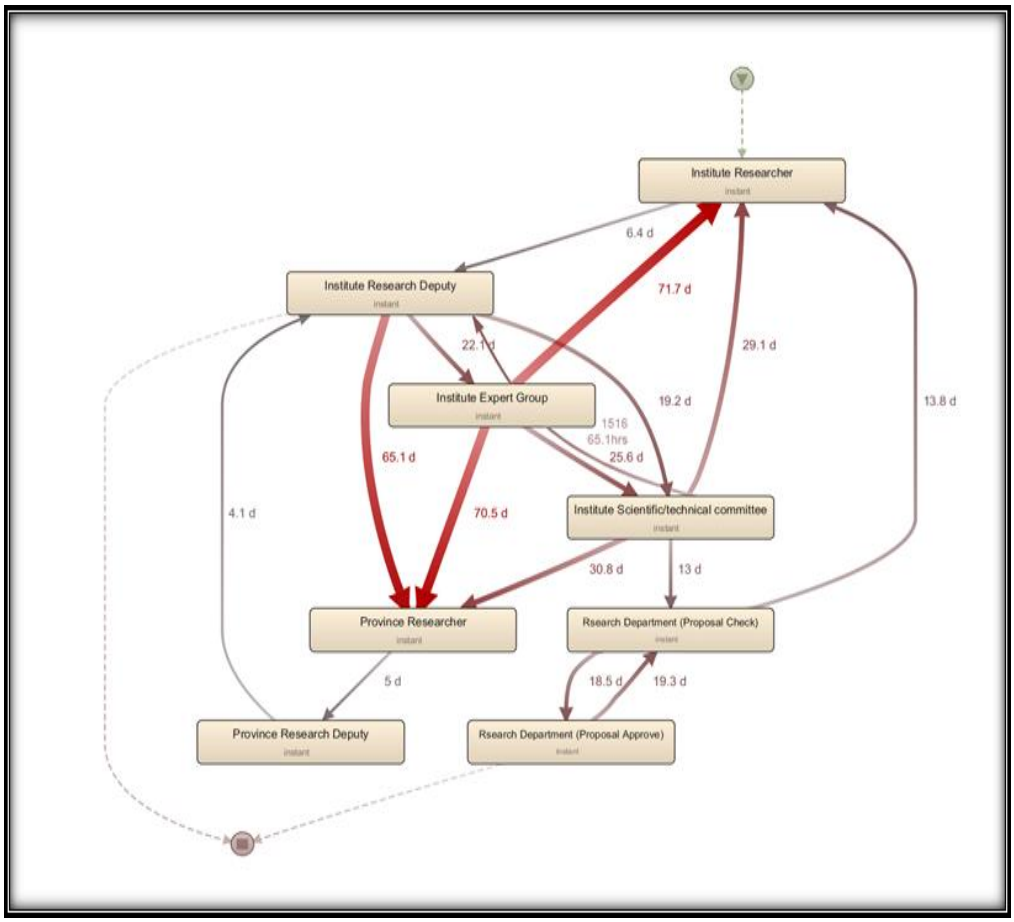


Figure 17: Time performance-mean duration time

The Figure 17 shows the min time 65.1 hours is for transaction of proposal survey and forwarding by Institute Scientific/technical committee to Institute Research Deputy. After that the mean time 4.1 days is for latency of transaction between Province Research Deputy and Institute Research Deputy. Communication between Province Researcher and Province Research Deputy has less delay in comparison with this similar communication in institute. The max time is related to transaction between Institute Expert Group and Province Researcher (70.5 days), Institute Research Deputy and Province Researcher (65.1 days) and Institute Scientific/technical committee and Province Researcher (30.8 days) respectively. This time in relation to institute Researcher is less significantly.

In our process model more than 1500 persons are involved. Participation rates from 4225 to 1 in reference number. The most frequency of reference to a person is 4225 that its relative frequency is 8.73%. This person has the role of Research Department (Proposal Check) in the process. The next frequency equals 3623 with relative frequency 7.49% and is related to the deputy of one of the most important institute in AREEO. And the third most referenced person is related to the role of Research Department (Proposal Approve).

Figure 18: Role relative frequency

Institute Name	Person with Role of	Frequency	Relative frequency
Headquarter Department	Research Department (Proposal Check)	4,225	8.73%
Seed and Plant Improvement Institute	Institute Research Deputy	3,623	7.49%
Headquarter Department	Research Department (Proposal Approve)	3,275	6.77%
Seed and Plant Improvement Institute	Institute Scientific/technical committee	3,217	6.65%
Cereal Research Group Seed and Plant Improvement Institute	Institute Expert Group	1,023	2.11%
Animal Science Research	Institute Research Deputy	975	2.02%
Agricultural Engineering Research	Institute Research Deputy	911	1.88%
Dryland Agricultural Research	Institute Research Deputy	779	1.61%
Agricultural Engineering Research	Institute Scientific/technical committee	768	1.59%
Animal Science Research	Institute Scientific/technical committee	666	1.38%
Plant Protection	Institute Research Deputy	640	1.32%
Plant Protection	Institute Scientific/technical committee	597	1.23%
Dryland Agricultural Research	Institute Scientific/technical committee	544	1.12%
Razi Vaccine and Serum Research	Institute Research Deputy	512	1.06%
Khorasan Razavi Province	Province Research Deputy	506	1.05%
Research Institute of Forest and Rangelands	Institute Research Deputy	486	1%

Considering this statistics and mean duration in every transaction committed by related role, our system has relative human resource efficiency.

Mining the Performance Perspective

The following chart shows the number of proposals and the related case duration. Table 5 includes the Chart statistics.

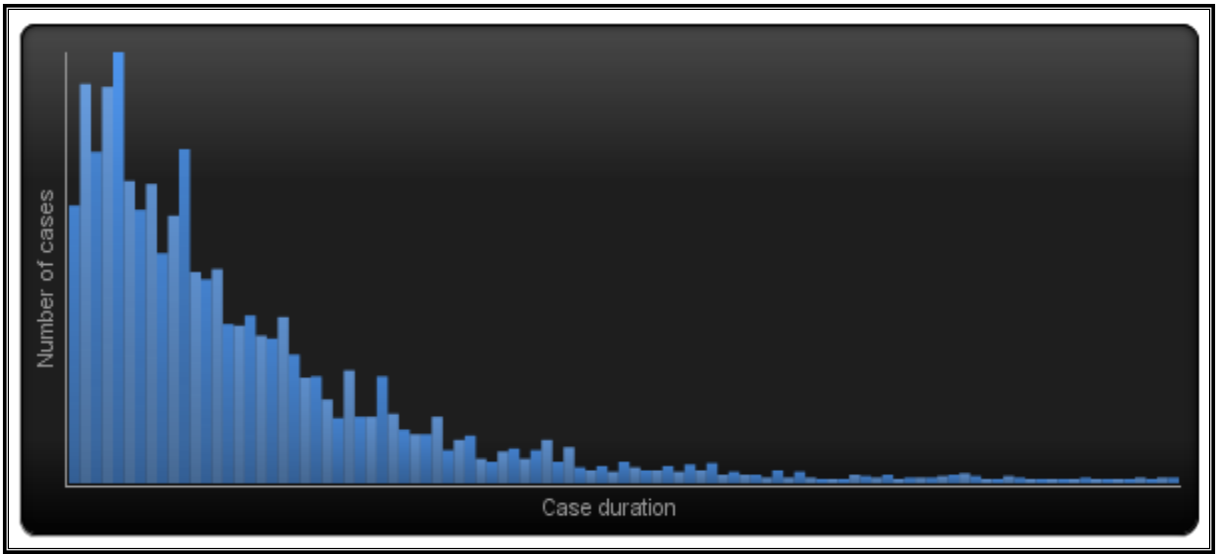


Figure 19: number of proposals/related case duration

The Figure 19 shows the relative time efficiency related to involved collection in the process. About 36% of proposals have been approved in the time less than 90 days. The minimum approve time is 11 days for 189 proposals. 28% of the proposals have passed the time between 3 to 6 months for approve. Approve time is 6 to 9 months for 16.3% of proposals. About 2% of proposals have had approve time 9 to 12 months and the 18% of proposals have waited for approve Unexplained time between one year to three years and three months. 82% of proposals have evaluation and approve time less than 1 year that considering every aspect such as evaluation time in Expert Group and Institute Scientific/technical committee, which both of them includes sending proposals for judgment, are reasonable and acceptable.

Figure 20: number of proposals/ Approve Time

Proposals	Approve Time	Time Classification	Frequency	Percent	The cumulative percentage
189	11 days	<90 days	1648	35.7	35.7
273	23days				
226	34days				
271	46days				
295	58days				
206	69 days				
186	81 days				
204	93 days	>90days and <180 days	1303	28.2	63.9
156	104 days				
182	116 days				
228	127 days				
143	139 days				
138	151 days				
145	162 days				
107	174 days				
106	186 days				
113	197 days	>180days and <270days	754	16.3	80.2
99	209 days				
97	221 days				

112	232 days				
86	244 days				
70	255 days				
71	267 days				
55	279 days	>270days and <1 year	408	1.9	82.1
42	290 days				
75	302 days				
43	314 days				
43	325 days				
71	337 days				
45	349 days				
34	360 days				
31	1 year and 7 days	>1 year and <1 year and 6 months	343	7.4	89.5
31	1 year and 18 days				
43	1 year and 30 days				
20	1 year and 42 days				
27	1 year and 53 days				
30	1 year and 65 days				
14	1 year and 77 days				
12	1 year and 88 days				
19	1 year and 100 days				
21	1 year and 112 days				
14	1 year and 123 days				
20	1 year and 135 days				
27	1 year and 146 days				
12	1 year and 158 days				
22	1 year and 170 days				
8	1 year and 181 days	>1 year and 6 months and <2 years	112	2.3	91.8
6	1 year and 193 days				
9	1 year and 205 days				
5	1 year and 216 days				
12	1 year and 229 days				
8	1 year and 240 days				
6	1 year and 251 days				
6	1 year and 263 days				
9	1 year and 274 days				
5	1 year and 286 days				
10	1 year and 298 days				
6	1 year and 309 days				
11	1 year and 321 days				
3	1 year and 333 days				
5	1 year and 344 days				
3	1 year and 356 days				
3	2 years and 3 days	>2 years and <3 years and 3 months	47	1	100
1	2 years and 14 days				
6	2 years and 26 days				
1	2 years and 37 days				
5	2 years and 49 days				

1	2 years and 61 days				
1	2 years and 69 days				
3	2 years and 107 days				
2	2 years and 119 days				
1	2 years and 131 days				
3	2 years and 142 days				
1	2 years and 177 days				
1	2 years and 189 days				
2	2 years and 200 days				
3	2 years and 212 days				
4	2 years and 224 days				
2	2 years and 270 days				
1	2 years and 282 days				
1	2 years and 352 days				
1	3 years and 45 days				
1	3 years and 80 days				
4613			4613		

In Time Efficiency, another notable factor is the maximum and minimum events per day. The Figure 20 shows distribution of transactions follow similar pattern in different years. Except the first and last month of year (the solar year), that researchers are more enthusiastic for proposing projects and pursuing proposed projects for approve them in current year or the beginning of New Year, in other days of year events number is between 20 and 90 per day. In these two periods the number of events per day reaches to 90 to 150.

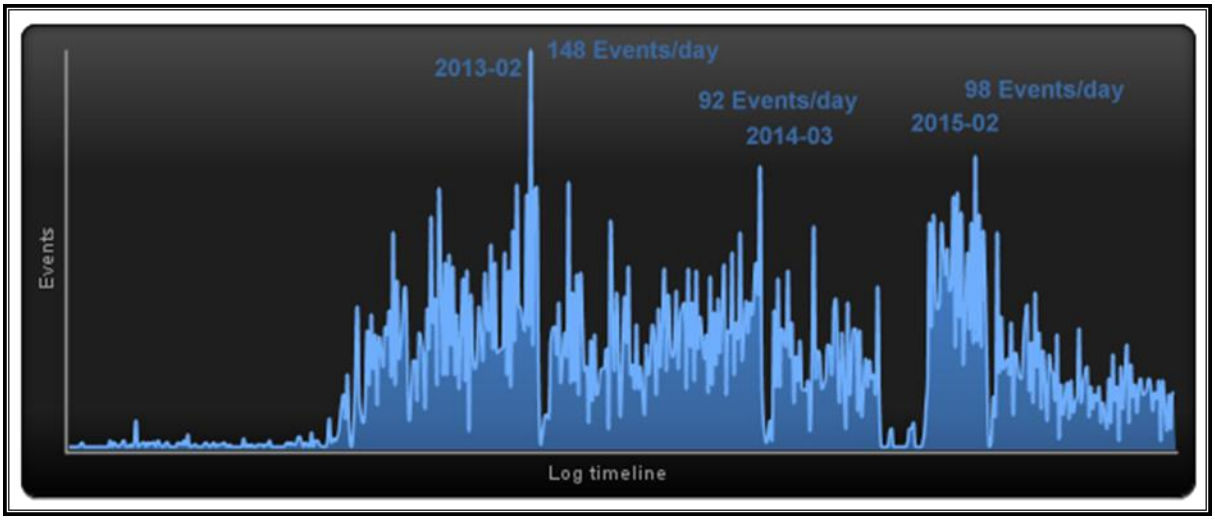


Figure 21: Events/ Log TimeLine

The following Figure depicts the distribution of proposals, which mentioned events in graph above occurred on them, over time. As showed in the Figure 19, this distribution in survey years in similar period of time is similar.

Similar to Figure 21, in the period between the beginning and end of solar year, the number of proposals, on which the transaction takes place, reaches 240 cases. But in the remaining days of the year this number is about 180 proposals.

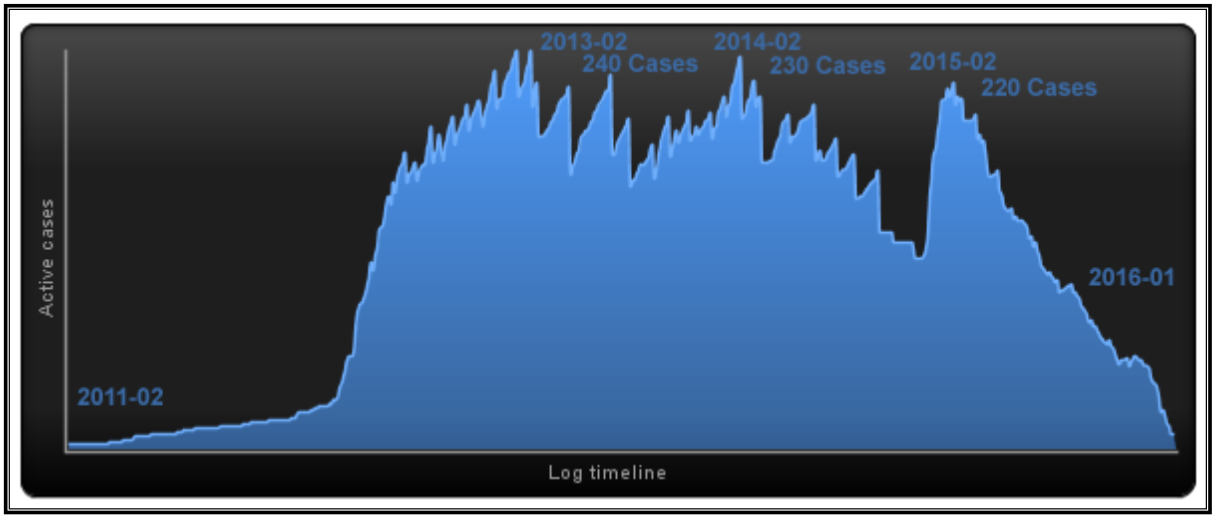


Figure 22: Active cases/ Log TimeLine

Mining the Case Perspective

As explained as follow in table 1, 1632 proposals passed the most repeated routes with the least steps which are 37% of all approved proposals. From all of proposals, 922 cases have their own exclusive path that shows the most flexibility in route selection (for consent improving of 25 primary and single product research Institutes with special discipline for evaluation and approve). From another point of view, this variety should be considered as critical subject. 1852 cases have passed routes repeated between 2 to 47 times. It is notable, at the first year of system operation; we had return ways, different of final selected route, leaded to sub direction been deactivated after a while. Although the most important change in the process was in 2013 that changed the final point of route from a headquarter department to institute research deputies, according to management policies. For this reason, it is viewable, in diagram 1 and 2 that there are two end points for proposal approve.

The path filtering feature in Disco helped to discover the routes with less than 10 repetitions and investigated that are they for existence of noisy data or not and this method resulted to edit or delete of them.

Figure 23: the most repeated routes with the least steps

Cases No	Step 1	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10	Step 11	Step 12	Step 13
538	Institute Researcher	Institute research deputy	Scientific/ Technical committee	Research Department (Check)	Research Department (Approve)	Approved								
349	Institute Researcher	Institute research deputy	Institute Expert Group	Scientific/ Technical committee	Research Department (Check)	Research Department (Approve)	Approved							

48	51	53	58	60	69	70	91	99	146
Institute Researcher	Institute Researcher	Province Researcher	Institute Researcher	Institute Researcher	Institute Researcher	Institute Researcher	Province Researcher	Institute Researcher	Institute Researcher
Institute research deputy	Institute research deputy	Province research deputy	Institute research deputy	Institute research deputy	Institute research deputy	Institute research deputy	Province research deputy	Institute research deputy	Institute research deputy
Institute Expert Group	Institute Expert Group	Institute research deputy	Institute Expert Group	Institute Expert Group	Scientific/ Technical committee	Institute Expert Group	Institute research deputy	Institute Expert Group	Scientific/ Technical committee
Institute Researcher	Scientific/ Technical committee	Institute Expert Group	Scientific/ Technical committee	Scientific/ Technical committee	Research Department (Check)	Institute Researcher	Institute Expert Group	Scientific/ Technical committee	Institute research deputy
Institute research deputy	Institute Expert Group	Scientific/ Technical committee	Institute Expert Group	Institute Expert Group	Scientific/ Technical committee	Institute Research deputy	Scientific/ Technical committee	Institute research deputy	Approved
Institute Expert Group	Institute Researcher	Institute Expert Group	Institute Researcher	Institute Researcher	Research Department (Check)	Institute research deputy	Research Department (Check)	Approved	
Scientific/ Technical committee	Institute research deputy	Province Researcher	Institute research deputy	Institute research deputy	Research Department (Approve)	Scientific/ Technical committee	Research Department (Approve)		
Institute research deputy	Institute Expert Group	Province research deputy	Scientific/ Technical committee	Institute Expert Group	Approved	Research Department (Check)	Approved		
Approved	Scientific/ Technical committee	Institute research deputy	Research Department (Check)	Scientific/ Technical committee		Research Department (Approve)			
	Institute research deputy	Institute Expert Group	Research Department (Approve)	Research Department (Check)		Approved			
	Approved	Scientific/ Technical committee	Approved	Research Department (Approve)					
		Research Department (Check)		Approved					
		Research Department (Approve)							
		Approved							

Another model efficiency parameter is the number of passed steps or transactions from the start to end of process. Figure 24 and Figure 25 show the relation between number of cases and events per case.

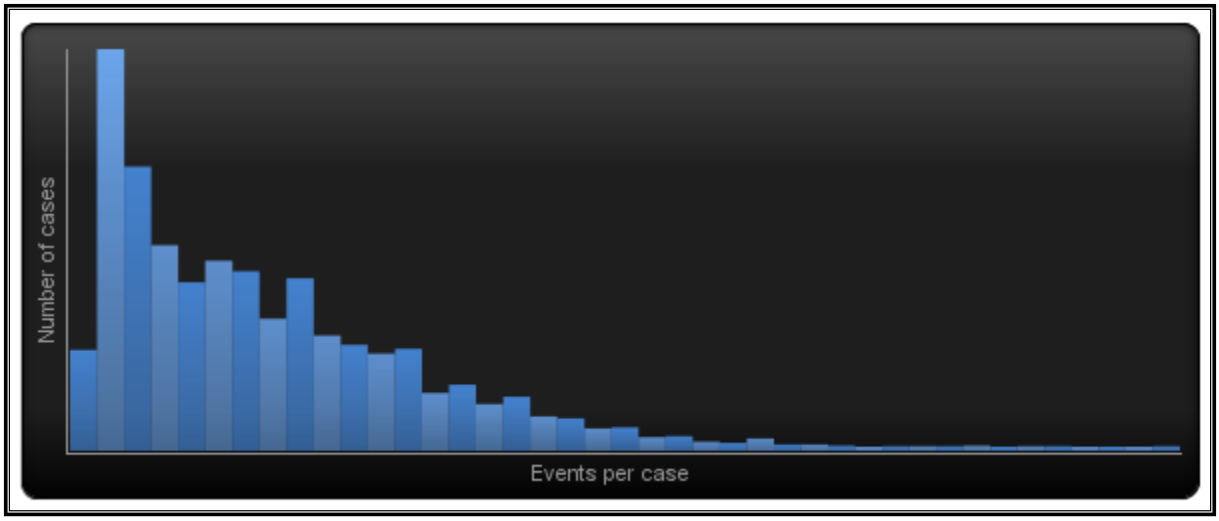


Figure 24: Number of Cases/Events per Case

Figure 25: Number of Cases/Events per Case

Cases	Steps	Time Classification	Frequency	Percent	The cumulative percentage
175	4 steps	<10	2713	58.9	58.9
718	5 steps				
506	6 steps				
364	7 steps				
297	8 steps				
336	9 steps				
317	10 steps				
231	11 steps	<20 and >10	1641	25.5	84.4
304	12steps				
201	13steps				
184	14steps				
168	15steps				
177	16steps				
97	17 steps				
112	18steps				
77	19 steps				
90	20steps				
55	21steps	<30 and >20	245	5.3	89.7
51	22steps				
33	23steps				
35	24steps				
17	25steps				
19	26steps				
9	27steps				
7	28steps				
15	29steps				
4	30steps				

4	31steps	<44 and >30	14	0.3	100
2	32steps				
1	34steps				
1	35steps				
1	36steps				
2	37 steps				
1	39steps				
1	40steps				
1	44steps				
			4613		

Figure 24 and Figure 25 also show that Process Model is successful about the number of necessary transactions for the navigation of process and reaching to end approve point. The number of anomalies, proposals have passed more than 30 steps in process unusually, is less than 0.3%.

Conclusion and Discussions

In this paper, the process surveyed of different dimensions. The goals are extracting knowledge which can be used as a tool for decision supporting of organization management. By mining in process perspective can result this system has so flexibility for research institute satisfaction. By considering the conclusions manager can decide limit these paths or not? Is this variety of paths acceptable or not?

The extracted model of process has enough clarity to show what happens in real and in spite of documents can be used for increasing of managers' awareness about system.

Mining the organization perspective clarify the key role of Institutes Research Deputy with the most hits in systems. In institutes, Seed and Plant Improvement Institute has the most efficiency, both in all key roles activity amount and response time, concurrently.

Mining the case perspective shows the institute proposals have notable priority in passing all steps of process in compare with province proposals. This can be explained by closely distance and face to face following. Another model efficiency parameter is the number of passed steps or transactions from the start to end of process. Our survey demonstrate the system has significant efficiency after omission of headquarters' Monitoring and Evaluation and assignment of final approve job to institutes.

References

[1] Agrawal, R., Gunopulos, D. & Leymann, F. 1998. 'Mining Process Models from Work-flow Logs'. In Sixth International Conference on Extending Database Technology, pp. 469-483.

[2] Gűnther, C. W. & van der Aalst, W. M. P. 2007. 'Fuzzy Mining – Adaptive Process Simplification Based on Multi-Perspective Metrics'. In: Alonso, G., Dadam, P., Rosemann, M. (eds), Business Process Management, 5th International Conference (BPM 2007), September 24-28 2007, Brisbane, Australia, Proceedings. Lecture Notes in Computer Science 4714, pp. 328-343, Springer.

[3] Jagadeesh Chandra Bose, R. P. & van der Aalst, W. M. P. 2009. 'Context Aware Trace Clustering: Towards Improving Process Mining Results'. Proceedings of the SIAM International Conference on Data Mining (SDM 2009), April 30 - May 2 2009. Sparks, Nevada, USA. pp. 401–412.

[4] Lemos, A. M., Sabino, C. C., Lima, R. M. F. & Oliveira, C. A. L. 2011. 'Using process mining in software development process management: A case study'. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC 2011), October 9-12 2011. Anchorage, Alaska, USA, pp. 1181-1186.

[5] Mans, R. S., Schonenberg, M. H., Song, M., van der Aalst, W. M. P. & Bakker, P. J. M. 2008. 'Process Mining in Healthcare - A Case Study'. In: In L. Azevedo and A.R. Londral (eds), Proceedings of the First International Conference on Health Informatics (HEALTHINF'08), January 28-31 2008. Funchal, Madeira, Portugal, Institute for Systems and Technologies of Information, Control and Communication, pp. 118-125, IEEE Computer Society.

[6] Maruster, L. &Beest, N. R. T. P. v. 2009. 'Redesigning business processes: a methodology based on simulation and process mining techniques'. Knowledge Information Systems 21, pp. 267-297.

[7] Rebuge, A. & Ferreira, D. R. 2012, 'Business process analysis in healthcare environments: A methodology based on process mining'. Information Systems 37(2), pp. 99-116.

[8] Reijers, H. A., Song, M. &Jeong, B. 2009. 'Analysis of a collaborative workflow process with distributed actors'. Information Systems Frontiers 11(3), pp. 307-322.
[9] Rozinat, A., Jong, I. S. M. d., Günther, C. W. & van der Aalst, W. M. P. 2009. 'Process Mining Applied to the Test Process of Wafer Scanners in ASML'. IEEE Transactions on Systems, Man, and Cybernetics, Part C (RSMC) 39, pp. 474-479.
[10] Rozinat, A. & van der Aalst, W. M. P. 2008. 'Conformance Checking of Processes Based on Monitoring Real Behavior'. Information Systems 33(1), pp. 64-95.
[11] Song, M. & van der Aalst, W. M. P. 2008. 'Towards Comprehensive Support for Organizational Mining'. Decision Support Systems 46(1), pp. 300-317.
[12] Song, M., Gunther, C. W. & van der Aalst, W. M. P. 2008. 'Trace Clustering in Process Mining'. In: Ardagna, D., Mecella, M., Yang, J. (eds), Business Process Management Workshops (BPM 2008), September 1-4 2008. Milano, Italy, Lecture Notes in Business Information Processing 17, pp. 109-120, Springer.
[13] Song, M., Yang, H., Siadat, S. H. & Pechenizkiy, M. 2013. 'A comparative study of dimensionality reduction techniques to enhance trace clustering performances', Expert Systems with Applications 40(9), pp. 3722-3737.
[14] Van der Aalst, W. M. P., Reijers, H. A., Weijters, A. J. M. M., van Dongen, B. F., de Medeiros, A. K. A., Song, M. & Verbeek, H. M. W. 2007. 'Business Process Mining: An Industrial Application'. Information Systems 32(5), pp. 713-732.
[15] Van der Aalst, W. M. P., van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G. & Weijters, A. J. M. M. 2003. 'Workflow Mining: A Survey of Issues and Approaches'. Data and Knowledge Engineering 47(2), pp. 237-267.
[16] Van der Aalst, W. M. P., Weijters, A. J. M. M. & Maruster, L. 2004. 'Workflow Mining: Discovering Process Models from Event Logs'. IEEE Transactions on Knowledge and Data Engineering 16(9), pp. 1128-1142.
[17] Martin Lang, Thomas B¨urkle, Susanne Laumann, and Hans-Ulrich Prokosch. Process mining for clinical workflows: Challenges and current limitations. In MIE, pages 229–234, 2008.
[18] Richard Lenz and Manfred Reichert. It support for healthcare processes - premises, challenges, perspectives. Data Knowl. Eng., 61(1):39–58, April 2007.
[19] M. Poulymenopoulou, F. Malamateniou, and G. Vassilacopoulos. Specifying workflow process requirements for an emergency medical service. J. Med. Syst., 27(4):325–335, August 2003.
[20] Wil M. P. van der Aalst, Arthur H. M. ter Hofstede, Bartek Kiepuszewski, and Alistair P. Barros. Workflow patterns. Distributed and Parallel Databases, 14(1):5–51, 2003.